

# Acoustic Source Separation of Convolutional Mixtures Based on Intensity Vector Statistics

Banu Günel, *Member, IEEE*, Hüseyin Hacihabiboğlu, *Member, IEEE*, and Ahmet M. Kondoç, *Member, IEEE*

**Abstract**—Various techniques have previously been proposed for the separation of convolutional mixtures. These techniques can be classified as stochastic, adaptive, and deterministic. Stochastic methods are computationally expensive since they require an iterative process for the calculation of the demixing filters based on a separation criterion that usually assumes that the source signals are statistically independent. Adaptive methods, such as the adaptive beamformers, also exploit signal properties in order to optimize a multichannel filter structure. However, these algorithms need initialization and time to converge. Deterministic methods, on the other hand, provide a closed-form solution based on the deterministic aspects of the problem, such as the channel characteristics and the source directions. This paper presents a technique that exploits the intensity vector statistics to achieve a nearly closed-form solution for the separation of the convolutional mixtures as recorded with a coincident microphone array. No assumptions are made on the signals, but it is assumed that the source directions are known *a priori*. Directivity functions based on von Mises functions are designed for beamforming depending on the circular statistics of the calculated intensity vectors. Numerical evaluation results were presented for various speech and instrument sounds and source positions in two reverberant rooms.

**Index Terms**—Array processing, B-format, beamforming, circular statistics, coincident microphone array, convolutional mixtures, intensity, source separation, von Mises.

## I. INTRODUCTION

THE separation of convolutional mixtures aims to estimate the individual sound signals in the presence of other such signals in reverberant environments. As sound mixtures are almost always convolutional in enclosures, their separation is a useful preprocessing stage for speech recognition and speaker identification problems. Other direct application areas also exist such as in hearing aids, teleconferencing, multichannel audio, and acoustical surveillance. Several techniques have been proposed before for the separation of convolutional mixtures, which can be grouped into three different categories: stochastic, adaptive, and deterministic.

Stochastic methods, such as the independent component analysis (ICA), are based on a separation criterion that assumes the statistical independence of the source signals [1], [2]. ICA was originally proposed for instantaneous mixtures. It is applied in

the frequency domain for convolutional mixtures, as the convolution corresponds to multiplication in the frequency domain. Although faster implementations exist such as the FastICA [3], [4], stochastic methods are usually computationally expensive due to the several iterations required for the computation of the demixing filters. Furthermore, frequency domain ICA-based techniques suffer from the scaling and permutation issues resulting from the independent application of the separation algorithms in each frequency bin [5], [6].

The second group of methods are based on adaptive algorithms that optimize a multichannel filter structure according to the signal properties. Depending on the type of the microphone array used, adaptive beamforming (ABF) utilizes spatial selectivity to improve the capture of the target source while suppressing the interferences from other sources [7], [8]. These adaptive algorithms are similar to stochastic methods in the sense that they both depend on the properties of the signals to reach a solution. It has been shown that the frequency domain adaptive beamforming is equivalent to the frequency domain blind source separation (BSS) [9]. However, the former assumes that the geometry of the array is known, while the latter does not utilize this information. The ABF algorithms need to adaptively converge to a solution which may be suboptimal. They also need to tackle with all the targets and interferences jointly. Furthermore, the null beamforming applied for the interference signal is not very effective under reverberant conditions due to the reflections, creating an upper bound for the performance of the BSS [10].

Deterministic methods, on the other hand, do not make any assumptions about the source signals and depend solely on the deterministic aspects of the problem such as the source directions and the multipath characteristics of the reverberant environment [11]–[13]. Although there have been efforts to exploit direction-of-arrival (DOA) information and the channel characteristics for solving the permutation problem [8], [14]–[17], this information was used in an indirect way, merely to assist the actual separation algorithm, which was usually stochastic or adaptive.

A deterministic approach that leads to a closed-form solution is very desirable from the computational point of view. However, no such method with satisfactory performance has been proposed so far. There are two reasons for this. First, the knowledge of the source directions is not sufficient for good separation, because without adaptive algorithms, the source directions can be exploited only by simple delay-and-sum beamformers. However, due to the limited number of microphones in an array, the spatial selectivity of such beamformers is not sufficient to perform well under reverberant conditions. Second, the

Manuscript received September 5, 2007; revised December 24, 2007. This work was supported by the Engineering and Physical Sciences Research Council (EPSRC) under Research Grant GR/S72320/01 Portfolio Partnership Award in Integrated Electronics. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Hiroshi Sawada.

The authors are with the Center for Communication Systems Research (CCSR), University of Surrey, Guildford, GU2 7XH, U.K. (e-mail: b.gunel@surrey.ac.uk; h.hacihabiboglu@ieee.org; a.kondoç@surrey.ac.uk).

Digital Object Identifier 10.1109/TASL.2008.918967

multipath characteristics of the environment cannot be found with sufficient accuracy while using noncoincident arrays, as the channel characteristics are different at each sensor position which in turn makes it difficult to determine the room responses from the mixtures.

Almost all of the source separation methods employ noncoincident microphone arrays to the extent that the existence of such an array geometry is an inherent assumption by default in the formulation of the problem. The use of a coincident microphone array was previously proposed to exploit the directivities of two closely positioned directional microphones [18]. However, the construction of the solution disregarded the fact that the reflections are weighted with different directivity factors according to their arrival directions for two directional microphones pointing at different angles. Therefore, the method was, in fact, not suitable for convolutive mixtures. In literature, coincident microphone arrays have been investigated mostly for intensity vector calculations and sound source localization [19]–[21].

This paper proposes a technique that provides a nearly closed-form solution for the separation of convolutive mixtures captured by a compact, coincident microphone array. The technique depends on the channel characterization in the frequency domain based on the analysis of the intensity vector statistics. Because of this reason, it does not suffer from the permutation problem which normally occurs due to inefficient channel modeling in the frequency domain methods. The directions of all sources and, therefore the number of sources, are assumed to be known or found at a preprocessing stage, such as with MULTIPLE Signal Classification (MUSIC) algorithm [22].

This paper is organized as follows. In Section II, the coincident array structure used in this work is introduced and the convolutive mixture signals captured by the coincident array geometry are formulated. Section III explains the theory and usage of circular intensity vector statistics for convolutive source separation as well as the implementation details at each processing stage. Section IV describes the experimental test conditions and provides the obtained results. Section V concludes the paper.

## II. FORMULATION OF MIXTURE SIGNALS FOR COINCIDENT ARRAYS

In the time–frequency domain, the pressure signal recorded by the  $m$ th microphone for  $N$  sources can be written as

$$p_m(\omega, t) = \sum_{n=1}^N h_{mn}(\omega, t) s_n(\omega, t) \quad (1)$$

where  $h_{mn}(\omega, t)$  is the time–frequency representation of the transfer function from the  $n$ th source to the  $m$ th microphone, and  $s_n(\omega, t)$  is the time–frequency representation of the  $n$ th original source. The aim of the sound source separation is estimating the mixture components from the observation of the microphone signals only.

Assume that four omnidirectional microphones are positioned very closely on a plane in the geometry as shown in Fig. 1. Each  $h_{mn}(\omega, t)$  coefficient can be represented as a plane wave arriving from direction  $\phi_n(\omega, t)$  with respect to the center

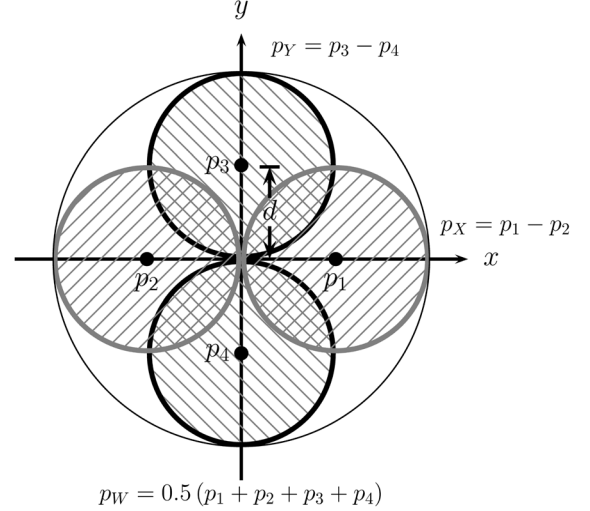


Fig. 1. Microphone array setup to obtain the pressure  $p_W$  and pressure gradients  $p_X$  and  $p_Y$  for intensity vector calculations. The  $p_i$  denotes the  $i$ th omnidirectional microphone.

of the array. Assuming the pressure at the center of the array due to this plane wave is  $p_o(\omega, t)$ . Then

$$h_{1n}(\omega, t) = p_o(\omega, t) e^{jkd \cos[\phi_n(\omega, t)]} \quad (2)$$

$$h_{2n}(\omega, t) = p_o(\omega, t) e^{-jkd \cos[\phi_n(\omega, t)]} \quad (3)$$

$$h_{3n}(\omega, t) = p_o(\omega, t) e^{jkd \sin[\phi_n(\omega, t)]} \quad (4)$$

$$h_{4n}(\omega, t) = p_o(\omega, t) e^{-jkd \sin[\phi_n(\omega, t)]} \quad (5)$$

where  $k$  is the wave number related to the wavelength  $\lambda$  as  $k = 2\pi/\lambda$ ,  $j$  is the imaginary unit and  $2d$  is the distance between the microphones. Now, define  $p_W = 0.5(p_1 + p_2 + p_3 + p_4)$ ,  $p_X = p_1 - p_2$  and  $p_Y = p_3 - p_4$ . Then

$$p_W(\omega, t) = \sum_{n=1}^N 0.5[h_{1n}(\omega, t) + h_{2n}(\omega, t) + h_{3n}(\omega, t) + h_{4n}(\omega, t)] s_n(\omega, t) \quad (6)$$

$$p_X(\omega, t) = \sum_{n=1}^N [h_{1n}(\omega, t) - h_{2n}(\omega, t)] s_n(\omega, t) \quad (7)$$

$$p_Y(\omega, t) = \sum_{n=1}^N [h_{3n}(\omega, t) - h_{4n}(\omega, t)] s_n(\omega, t). \quad (8)$$

If  $kd \ll 1$ , i.e., when the microphones are positioned close to each other in comparison to the wavelength, it can be shown by using the relations  $\cos(kd \cos \theta) \approx 1$ ,  $\cos(kd \sin \theta) \approx 1$ ,  $\sin(kd \cos \theta) \approx kd \cos \theta$  and  $\sin(kd \sin \theta) \approx kd \sin \theta$  that

$$p_W(\omega, t) \simeq \sum_{n=1}^N 2p_o(\omega, t) s_n(\omega, t) \quad (9)$$

$$p_X(\omega, t) \simeq \sum_{n=1}^N j2p_o(\omega, t) kd \cos[\phi_n(\omega, t)] s_n(\omega, t) \quad (10)$$

$$p_Y(\omega, t) \simeq \sum_{n=1}^N j2p_o(\omega, t)kd \sin[\phi_n(\omega, t)]s_n(\omega, t). \quad (11)$$

The  $p_W$  is similar to an omnidirectional microphone, and  $p_X$  and  $p_Y$  are similar to two bidirectional microphones that approximate pressure gradients along the  $X$  and  $Y$  directions, respectively. These signals are also known as B-format signals which can also be obtained by four capsules positioned at the sides of a tetrahedron [23]. The next section explains the use of these signals for source separation based on intensity vector analysis.

### III. SEPARATION USING INTENSITY VECTOR STATISTICS

#### A. Calculation of the Intensity Vectors

The acoustic particle velocity  $\mathbf{v}(\mathbf{r}, \omega, t)$  is defined in two dimensions as [24]

$$\mathbf{v}(\mathbf{r}, \omega, t) = \frac{1}{\rho_0 c} [p_X(\omega, t)\mathbf{u}_x + p_Y(\omega, t)\mathbf{u}_y] \quad (12)$$

where  $\rho_0$  is the ambient density,  $c$  is the speed of sound,  $\mathbf{u}_x$ , and  $\mathbf{u}_y$  are unit vectors in the directions of corresponding axes.

The product of the pressure and the particle velocity gives instantaneous intensity. The active intensity can be found as [24]

$$\mathbf{I}(\omega, t) = \frac{1}{\rho_0 c} [\text{Re}\{p_W^*(\omega, t)p_X(\omega, t)\}\mathbf{u}_x + \text{Re}\{p_W^*(\omega, t)p_Y(\omega, t)\}\mathbf{u}_y] \quad (13)$$

where  $*$  denotes conjugation and  $\text{Re}\{\bullet\}$  denotes taking the real part of the argument.

Then, the direction of the intensity vector  $\gamma(\omega, t)$  can be obtained by

$$\gamma(\omega, t) = \arctan \left[ \frac{\text{Re}\{p_W^*(\omega, t)p_Y(\omega, t)\}}{\text{Re}\{p_W^*(\omega, t)p_X(\omega, t)\}} \right]. \quad (14)$$

The reverberant estimate of the  $n$ th source,  $\tilde{s}_n$  is obtained by beamforming in the source direction with a directivity function  $J_n(\theta; \omega, t)$  so that

$$\tilde{s}_n(\omega, t) = p_W(\omega, t) J_n(\gamma(\omega, t); \omega, t). \quad (15)$$

By this weighting, the time–frequency components of the omnidirectional microphone signal are amplified more if the direction of the corresponding intensity vector is closer to the direction of the target source. It should be noted that this weighting also has the effect of partial deconvolution as the reflections are also suppressed depending on their arrival directions. The following section explains the calculation of the directivity function from the intensity vector statistics.

#### B. Determining the Directivity Function

The directivity function  $J_n(\theta; \omega, t)$  used for the  $n$ th source is a function of  $\theta$  only in the analyzed time–frequency bin. It is determined by the local statistics of the calculated intensity

vector directions  $\gamma(\omega, t)$  for the analyzed short-time window. It may be suggested that these directions are von Mises distributed due to reverberation.

In circular statistics, von Mises distribution is generally preferred to model circular data as this distribution can also approximate other circular distributions such as uniform, cardioid, wrapped normal and Cauchy [25].

The probability density function of von Mises distribution is given as

$$f(\theta; \mu, \kappa) = \frac{e^{\kappa \cos(\theta - \mu)}}{2\pi I_0(\kappa)} \quad (16)$$

for a circular random variable  $\theta$  where  $0 < \theta \leq 2\pi$ ,  $0 \leq \mu < 2\pi$  is the mean direction,  $\kappa > 0$  is the concentration parameter, and  $I_0(\kappa)$  is the modified Bessel function of order zero.

For  $N$  sound sources, the probability density function of the intensity vector directions can be modeled as a mixture of  $N$  von Mises probability density functions each with a mean direction of  $\mu_n$ , corresponding to the source directions, and a circular uniform density due to the isotropic late reverberation.

$$g(\theta) = \frac{\alpha_0}{2\pi} + \sum_{n=1}^N \alpha_n f(\theta; \mu_n, \kappa_n) \quad (17)$$

where,  $0 \leq \alpha_i \leq 1$  are the component weights, and  $\sum_i \alpha_i = 1$ .

As analytical methods do not exist for finding the maximum likelihood estimates of the mixture parameters, it can be assumed that the  $\alpha_n$  and  $\kappa_n$  take discrete values within some boundary and the values of these parameters that maximize the likelihood can be determined numerically. The directivity function for beamforming in the direction of the  $n$ th source for a given time–frequency bin is then defined as

$$J_n(\theta; \omega, t) = \alpha_n \frac{e^{\kappa_n(t) \cos(\theta - \mu_n)}}{2\pi I_0(\kappa_n(t))}. \quad (18)$$

Assuming all sound sources and late reverberation are present in the environment, an intensity vector can belong to any one of the sound sources or the late reverberation. The weight of the uniformly distributed late reverberation does not affect much the outcome of the numerical maximum likelihood estimation algorithm. Therefore, for simplicity, the component weights can be assumed to be equal to each other, i.e.,  $\alpha_n = 1/(N + 1)$ .

It can be shown by using the definition of the von Mises function in (16) that the concentration parameter  $\kappa$  is logarithmically related to the 6-dB beamwidth  $\theta_{BW}$  of this directivity function as

$$\kappa = \ln 2 / [1 - \cos(\theta_{BW}/2)]. \quad (19)$$

Then, in numerical maximum likelihood estimation, it is appropriate to determine the concentration parameters from linearly increasing beamwidth values. Fig. 2 shows four von Mises functions for 6-dB beamwidths of  $10^\circ$  ( $\kappa = 262.79$ ),  $45^\circ$  ( $\kappa = 13.14$ ),  $90^\circ$  ( $\kappa = 3.41$ ) and  $180^\circ$  ( $\kappa = 1.00$ ).

It should be noted that while (15) suggests a closed-form solution, determining the  $\kappa$  parameters requires parameter searches within all possible solutions. Therefore, the solution is only nearly closed-form.

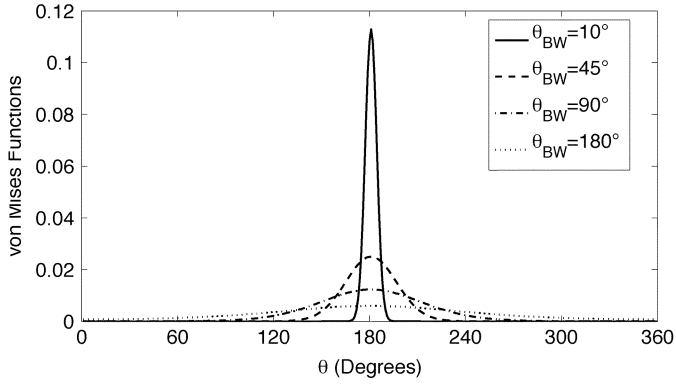


Fig. 2. The von Mises functions for 6-dB beamwidths of  $10^\circ$ ,  $45^\circ$ ,  $90^\circ$ , and  $180^\circ$ , corresponding the concentration parameters,  $\kappa$  of 262.79, 13.14, 3.41, and 1.00, respectively.

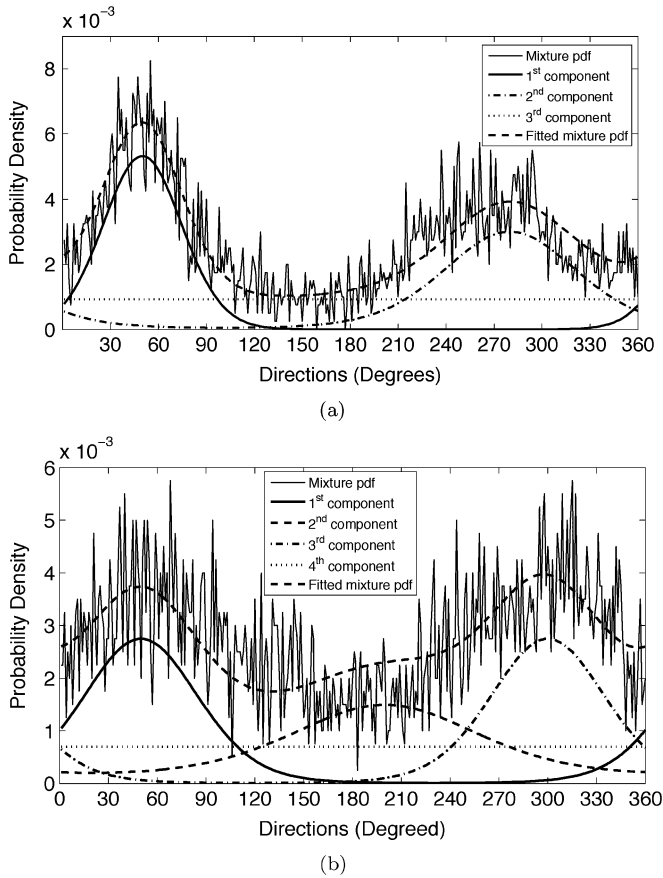


Fig. 3. Probability density function of the intensity vector directions, individual mixture components, and fitted mixtures (a) for two sources at  $50^\circ$  and  $280^\circ$  and (b) for three sources at  $50^\circ$ ,  $200^\circ$ , and  $300^\circ$ .

Fig. 3(a) and (b) show the probability density functions of the intensity vector directions, individual mixture components, and the fitted mixtures for two and three speech sources, respectively. The sources are at  $50^\circ$  and  $280^\circ$  for the first figure and  $50^\circ$ ,  $200^\circ$ , and  $300^\circ$  for the second figure. The intensity vector directions were calculated for an exemplary analysis window of length 4096 samples at 44.1 kHz in a room with reverberation time of 0.83 s.

It should be noted that the fitting is applied to determine the directivity functions. Therefore, testing the goodness-of-fit by

methods such as the Kuiper test [25] is irrelevant. The geometrical positioning of the sources and the microphones may also have an effect on the fitting performance and consequently the separation performance. For example, when sound sources are close to strong reflecting surfaces the observed source direction would shift away from the actual direction. In such a situation, the usage of the observed source direction instead of the actual source direction may be considered.

### C. Implementation

The processing stages of the proposed technique can be divided into five steps as shown in Fig. 4:

- Step 1) The pressure and pressure gradient signals are calculated from the array described in Section II, or obtained directly in B-format by using one of the commercially available tetrahedron microphones. The spacing between the microphones should be small to avoid aliasing at high frequencies. Phase errors at low frequencies should also be taken into account if a reliable frequency range for operation is essential [24].
- Step 2) Time–frequency representations of the pressure and pressure gradient signals are calculated using the modified discrete cosine transform (MDCT) where subsequent blocks are overlapped by 50% [26]. The MDCT is chosen due to its overlapping and energy compaction properties to decrease the edge effects across blocks that occur as the directivity function used for each time–frequency bin changes. Similar to discrete Fourier transform (DFT), MDCT has the convolution-multiplication property [27]. Although the relationship between the time-domain and transform domain signals is not as straightforward as in the case of DFT, this does not have any effect on the algorithm, because the weighting functions are found according to the transform domain properties of the signals and used in this domain before reconstruction. Perfect reconstruction is achieved with a window function  $w_k$  that satisfies  $w_k^2 + w_{k+M}^2 = 1$ , where  $2M$  is the window length [28]. In this paper, the following window function is used:

$$w_k = \sin\left(\frac{\pi}{2} \sin^2\left[\frac{\pi}{2M}\left(k + \frac{1}{2}\right)\right]\right). \quad (20)$$

- Step 3) The intensity vector directions are calculated and rounded to the nearest degree. The mixture probability density is obtained from the histogram of the found directions. Then, the statistics of these directions are analyzed in order to estimate the mixture component parameters as in (17). For numerical maximum-likelihood estimation, the 6-dB beamwidth is spanned linearly from  $10^\circ$  to  $180^\circ$  with  $10^\circ$  intervals and the related concentration parameters are calculated by using (19). Beamwidths smaller than  $10^\circ$  were not included since very sharp clustering around a source direction was not observed from the densities of the intensity vector

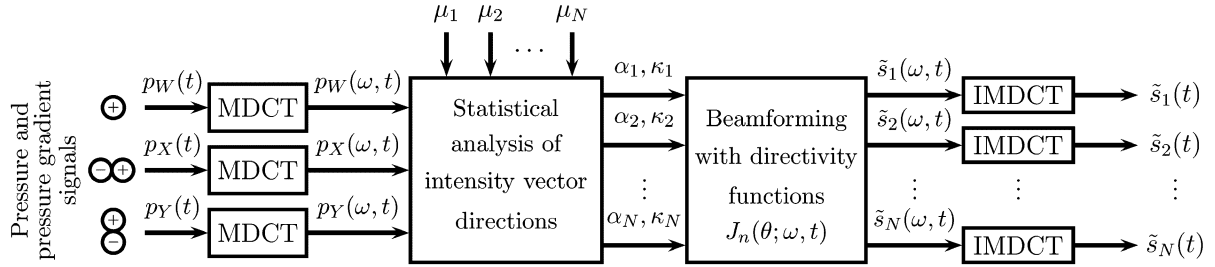


Fig. 4. Processing stages of the implementation. Initially, the pressure and pressure gradient signals are obtained from the array. Then, the modified discrete cosine transform (MDCT) of these signals are calculated. Next, the intensity vector directions are calculated and using the known source directions, von Mises mixture parameters are estimated. Next, beamforming is applied for each of the target sources using the directivity functions obtained from the von Mises functions. Finally, inverse modified cosine transform (IMDCT) of the separated signals are calculated, which reveals the time-domain estimates of the sound sources.

directions. As the point source assumption does not hold for real sound sources, such clustering is not expected even in anechoic environments due to the observed finite aperture of a sound source at the recording position. Beamwidths more than  $180^\circ$  were also not considered as the resulting von Mises functions are not very much different from the uniform density functions.

- Step 4) A directivity function is defined for each sound source and time–frequency bin as in (18), and beamforming is applied as in (15). Theoretically, an infinite number of directivity functions can be calculated to separate more sources than microphones, although the performance would be limited. For large number of sources, it may be more practical to use fixed directivity functions for each window, as their calculation would be computationally demanding.
- Step 5) Finally, the inverse modified cosine transform (IMDCT) of each separated signal is calculated to obtain the separated signals in the time-domain.

#### IV. EXPERIMENTAL RESULTS

The proposed algorithm was tested for mixtures of two and three sources for various source positions, in two rooms with different reverberation times. The recording setup, procedure for obtaining the mixtures, and the performance measures are discussed first, followed by the results presenting various factors that affect the separation performance.

##### A. Obtaining the Mixtures

The convolutive mixtures used in the testing of the algorithm were obtained by first measuring the B-format room impulse responses, convolving anechoic sound sources with these impulse responses and summing the resulting reverberant recordings. This method exploits the linearity and time-invariance assumptions of the linear acoustics.

The impulse responses were measured in two different rooms. The first room was an ITU-R BS1116 standard listening room with a reverberation time of 0.32 s. The second one was a meeting room with a reverberation time of 0.83 s. Both rooms were geometrically similar ( $L = 8$  m;  $W = 5.5$  m;  $H = 3$  m) and were empty during the tests.

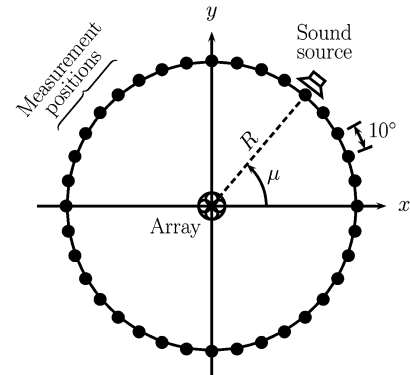


Fig. 5. Setup and measurement position of 36 room impulse responses positioned around a circle of radius  $R$  with  $10^\circ$  intervals.

For both rooms, 36 B-format impulse response recordings were obtained at 44.1 kHz with a SoundField microphone system (SPS422B) and a loudspeaker (Genelec 1030A), using a 16th-order maximum length sequence (MLS) signal [29]. Each of the 36 measurement positions were located on a circle of 1.6 m radius for the first room, and 2.0 m radius for the second room, as shown in Fig. 5. The recording points were at the center of the circles, and the frontal directions of the recording setup were fixed in each room. Source locations were selected between  $0^\circ$  to  $350^\circ$  with  $10^\circ$  intervals with respect to the recording setup. At each measurement position, the acoustical axis of the loudspeaker was facing towards the array location, while the orientation of the microphone system was kept fixed. The source and recording positions were 1.2 m high above the floor. The loudspeaker had a width of 20 cm, corresponding to the observed source apertures of  $7.15^\circ$  and  $5.72^\circ$  at the recording positions for the first and second rooms, respectively.

Anechoic sources sampled at 44.1 kHz were used from the Music for Archimedes CD [30]. The 5-s-long portions of male English speech (M), female English speech (F), male Danish speech (D), cello music (C), and guitar music (G) sounds were first equalized for energy, then convolved with the B-format impulse responses of the desired directions. The B-format sounds were then summed to obtain FM, CG, FC, and MG for two source mixtures and FMD, CFG, MFC, DGM for three source mixtures.

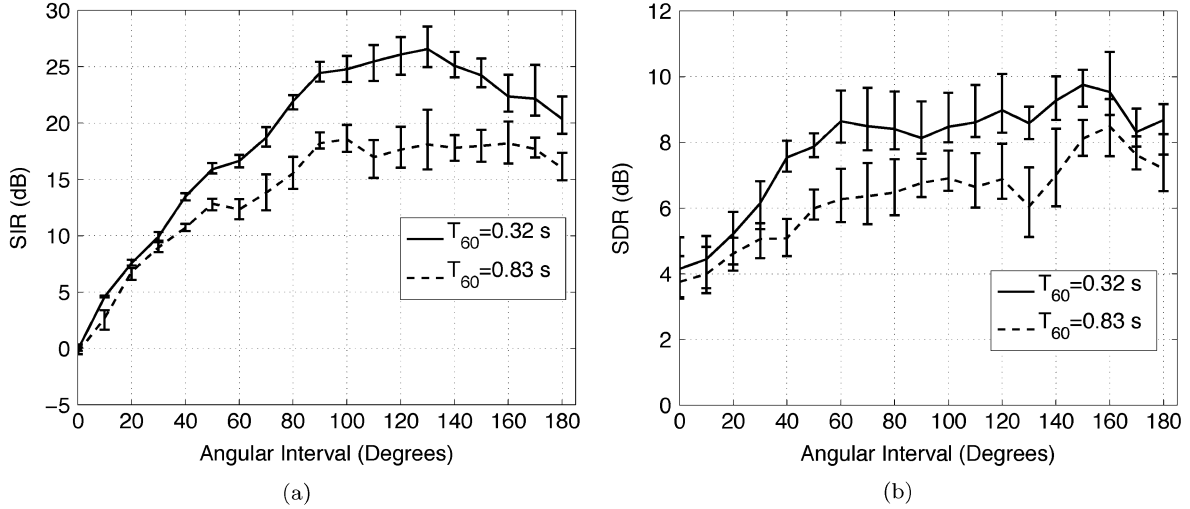


Fig. 6. (a) SIR and (b) SDR in dB with respect to the angular interval between the two sound sources for two rooms with reverberation times of 0.32 s and 0.83 s. The position of the first source of each mixture (FM, CG, FC, and MG) was fixed at  $0^\circ$ , while the position of the second source varied between  $0^\circ$  and  $180^\circ$  with  $10^\circ$  intervals.

### B. Performance Measures

There exist various criteria for the performance measure of source separation techniques. In this work, one-at-a-time signal-to-interference ratio (SIR) is used for quantifying the separation, as separately synthesized sources are summed together to obtain the mixture [31]. This metric is defined as

$$\text{SIR} = \frac{1}{N} \sum_{i=1}^N 10 \log \left[ \frac{E\{(\tilde{s}_{i|s_i})^2\}}{E\{(\sum_{i \neq j} \tilde{s}_{i|s_j})^2\}} \right] \quad (21)$$

where  $N$  is the total number of sources,  $\tilde{s}_{i|s_i}$  is the estimated source  $\tilde{s}_i$  when only source  $s_i$  is active,  $\tilde{s}_{i|s_j}$  is the estimated source  $\tilde{s}_i$  when only source  $s_j$  is active, and  $E\{\bullet\}$  is the expectation operator. Hild II *et al.* [32] has suggested for convolutive mixtures that the values of SIR above 15 dB indicates a good separation.

In addition to SIR, signal-to-distortion ratio (SDR) has also been used in order to quantify the quality of the separated sources [31]. However, the SDR is sensitive to the reverberation content of the original source used as the reference. If the anechoic source is used for comparison, this measure penalizes the effect of the reverberation even if the separation is quite good. On the other hand, if the reverberant source as observed at the recording position is used, then any deconvolution achieved in addition to the separation is also penalized as distortion.

When only one sound source is active, any of the B-format signals or cardioid microphone signals that can be obtained from them [33] can be used as the reference of that source. All of these signals can be said to have *perfect* sound quality, as the reverberation is not distortion. Therefore, it is fair to choose the reference signal that results in the best SDR values.

A hypercardioid microphone has the highest directional selectivity that can be obtained by using B-format signals providing the best signal-to-reverberation gain [33]. Since, the proposed technique performs partial deconvolution in addition to reverberation, a hypercardioid microphone most sensitive in

the direction of the  $i$ th sound source is synthesized from the B-format recordings when only one source is active, such that

$$p_{C_i|s_i} = \frac{1}{4} p_{W_i|s_i} + \frac{3}{4} (p_{X_i|s_i} \cos \mu_i + p_{Y_i|s_i} \sin \mu_i). \quad (22)$$

The source signal obtained in this way is used as the reference signal in the SDR calculation

$$\text{SDR} = \frac{1}{N} \sum_{i=1}^N 10 \log \left( \frac{E\{(\tilde{s}_i)^2\}}{E\{(\tilde{s}_i - \alpha_i p_{C_i|s_i})^2\}} \right) \quad (23)$$

where  $\alpha_i = E\{(\tilde{s}_i)^2\} / E\{(p_{C_i|s_i})^2\}$ .

### C. Mixtures of Two Sources

Fig. 6(a) and (b) show the SIR and SDR in dB plotted against the angular interval between the two sound sources. The first sound source was positioned at  $0^\circ$  and the position of the second source was varied from  $0^\circ$  to  $180^\circ$  with  $10^\circ$  intervals to yield the corresponding angular interval. The tests were repeated both for the listening room and for the meeting room. The error bars were calculated using the lowest and highest deviations from the mean values considering all four mixtures (FM, CG, FC, and MG).

As expected, better separation is achieved in the listening room than in the meeting room. The SIR values increase, in general, when the angular interval between the sound sources increases, although at around  $180^\circ$ , the SIR values decrease slightly because for this angle both sources lie on the same axis causing vulnerability to phase errors.

The SDR values also increase when the angular interval between the two sources increases. Similar to the SIR values, the SDR values are better for the listening room which has the lower reverberation time. The similar trend observed for the SDR and SIR values indicates that the distortion is mostly due to the interferences rather than the processing artifacts.

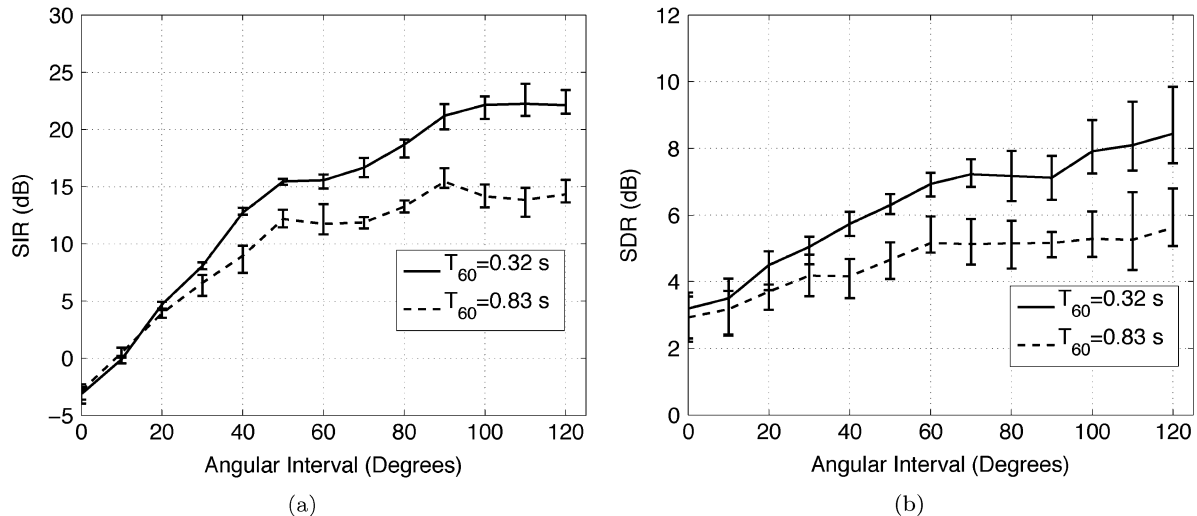


Fig. 7. (a) SIR and (b) SDR ratios in dB with respect to the angular interval between the three sound sources for two rooms with reverberation times of 0.32 and 0.83 s. The position of the first source of each mixture (FMD, CFG, MFC, and DGM) was fixed at  $0^\circ$ , while the position of the second and third sources varied to yield equal angular intervals between the sources.

TABLE I  
TEST CONDITIONS AND PERFORMANCES FOR SOME SOURCE SEPARATION TECHNIQUES

	T-F Masking [34]	GSS [8]	Fast Conv. BSS [17]	FD ICA Comp. Gr. [35]
Number of Sources	3	2	2	3
Number of Mics.	2	4	2	4
Reverberation Time	500 ms	50 ms	200 ms	130 ms
Angular Separation	$90^\circ, 90^\circ$	$45^\circ$	$60^\circ - 130^\circ$	$135^\circ, 165^\circ$
SIR Imp. (Approx.)	8.45 dB	15.00 dB	12.50 dB	16.44 dB

The sound sources were positioned at a longer distance in the meeting room than in the listening room, although the angular separations were the same. In general, the effects of room acoustics (i.e., early reflections and late reverberation) would be felt more pronounced when the sources are further away from the recording position. Therefore, different source distances might also have contributed to the performance differences in the two rooms, leading to worse separation in the meeting room.

#### D. Mixtures of Three Sources

Fig. 7(a) and (b) show the SIR and SDR in dB plotted against the angular interval between the three sound sources. The first sound source was positioned at  $0^\circ$ , the position of the second source was varied from  $0^\circ$  to  $120^\circ$  with  $10^\circ$  increasing intervals, and the position of the third source was varied from  $360^\circ$  to  $240^\circ$  with  $10^\circ$  decreasing intervals to yield the corresponding equal angular intervals from the first source. The tests were repeated both for the listening room and the meeting room. The error bars were calculated using the lowest and highest deviations from the mean values considering all four mixtures (FMD, CFG, MFC, and DMG).

The SIR values display a similar trend to the two-source mixtures, increasing with increasing angular intervals and taking higher values in the room with less reverberation time. The values, however, are lower in general from those obtained for the two-source mixtures, as expected.

The SDR values indicate better sound quality for larger angular intervals between the sources and for the room with less reverberation time. However, the quality is usually less than that obtained for the two-source mixtures.

#### E. Comparison With Other Methods

The performance of the convolutive source separation methods depends on many factors, such as the number of microphones, number of sound sources, the reverberation time of the rooms, and the positioning of the sources and microphones. Due to these reasons, a direct comparison cannot be made between the proposed method and the others. Table I summarizes real test conditions and performances of different techniques, namely, time-frequency masking [34], geometric source separation (GSS) [8], fast convergence BSS [17], and frequency-domain ICA with grouping separated frequency components via estimation of propagation model parameters [35].

As can be observed from Figs. 6(a) and 7(a) for similar test conditions, the performance of the proposed method is comparable to or better than some of the well-known source separation techniques.

## V. CONCLUSION

An acoustic source separation method for convolutive mixtures has been presented in this paper. It was shown that the intensity vector directions can be found by using the pressure and pressure gradient signals obtained from a closely spaced microphone array. The method assumes *a priori* knowledge of the sound source directions. The densities of the observed intensity vector directions are modeled as mixtures of von Mises density functions with mean values around the source directions and a uniform density function corresponding to the isotropic late reverberation. The statistics of the mixture components are then

exploited for separating the mixture by beamforming in the directions of the sources in the time–frequency domain.

The method was extensively tested for two and three source mixtures of speech and instrument sounds, for various angular intervals between the sources, and for two rooms with different reverberation times. The proposed method provides good separation as quantified by the SIR and SDR measures. The effects of angular interval, number of sources in the mixture, and reverberation time on the separation and distortion were discussed. The method performs better when the angular interval between the sources is large. Similarly, the method performs slightly better for the two-source mixtures in comparison with three-source mixtures. As expected, higher reverberation time reduces the separation performance and increases distortion.

Important advantages of the proposed method are the compactness of the array, low number of individual channels to be processed, and the simple nearly closed-form solution it provides as opposed to adaptive or iterative source separation algorithms. As such, the method can be used in teleconferencing applications, hearing aids, acoustical surveillance, and speech recognition among others.

#### REFERENCES

- [1] P. Comon, “Independent component analysis, a new concept?,” *Signal Process.*, vol. 36, no. 3, pp. 287–314, Apr. 1994.
- [2] J.-F. Cardoso, “Blind source separation: Statistical principles,” *Proc. IEEE*, vol. 86, no. 10, pp. 2009–2025, Oct. 1998.
- [3] A. Hyvärinen, “Fast and robust fixed-point algorithms for independent component analysis,” *IEEE Trans. Neural Netw.*, vol. 10, no. 3, pp. 626–634, May 1999.
- [4] S. C. Douglas, M. Gupta, H. Sawada, and S. Makino, “Spatio-temporal FastICA algorithms for the blind separation of convolutive mixtures,” *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 15, no. 5, pp. 1511–1520, Jul. 2007.
- [5] P. Smaragdis, “Blind separation of convolved mixtures in the frequency domain,” *Neurocomputing*, vol. 22, no. 1–3, pp. 21–34, 1998.
- [6] N. Mitianoudis and M. E. Davies, “Audio source separation: Solutions and problems,” *Int. J. Adapt. Control Signal Process.*, vol. 18, no. 3, pp. 299–314, Apr. 2004.
- [7] L. J. Griffiths and C. W. Jim, “An alternative approach to linearly constrained adaptive beamforming,” *IEEE Trans. Antennas Propag.*, vol. 30, no. 1, pp. 27–34, Jan. 1982.
- [8] L. C. Parra and C. V. Alvino, “Geometric source separation: Merging convolutive source separation with geometric beamforming,” *IEEE Trans. Speech Audio Process.*, vol. 10, no. 6, pp. 352–362, Sep. 2002.
- [9] S. Araki, S. Makino, Y. Hinamoto, R. Mukai, T. Nishikawa, and H. Saruwatari, “Equivalence between frequency-domain blind source separation and frequency-domain adaptive beamforming for convolutive mixtures,” *EURASIP J. Appl. Signal Process.*, vol. 2003, no. 11, pp. 1157–1166, 2003.
- [10] S. Araki, R. Mukai, S. Makino, T. Nishikawa, and H. Saruwatari, “The fundamental limitation of frequency domain blind source separation for convolutive mixtures of speech,” *IEEE Trans. Speech Audio Process.*, vol. 11, no. 2, pp. 109–116, Mar. 2003.
- [11] A.-J. van der Veen, “Algebraic methods for deterministic blind beamforming,” *Proc. IEEE*, vol. 86, no. 10, pp. 1987–2008, Oct. 1998.
- [12] H. Shindo and Y. Hirai, “Blind source separation by a geometrical method,” in *Proc. IEEE Int. Joint Conf. Neural Netw.*, Honolulu, HI, May 2002, vol. 2, pp. 1109–1114.
- [13] J. Yamashita, S. Tatsuta, and Y. Hirai, “Estimation of propagation delays using orientation histograms for anechoic blind source separation,” in *Proc. IEEE Int. Joint Conf. Neural Netw.*, Budapest, Hungary, Jul. 2004, vol. 3, pp. 2175–2180.
- [14] W. Wang, J. A. Chambers, and S. Sanei, “A novel hybrid approach to the permutation problem of frequency domain blind source separation,” in *Proc. Int. Conf. Independent Compon. Anal. Blind Signal Separation*, Granada, Spain, Sep. 2004, pp. 532–539.
- [15] N. Mitianoudis and M. Davies, “Permutation alignment for frequency domain ICA using subspace beamforming method,” in *Proc. Int. Conf. Independent Compon. Analysis Blind Signal Separation*, Granada, Spain, Sep. 2004, pp. 669–676.

- [16] M. Z. Ikram and D. R. Morgan, “A beamforming approach to permutation alignment for multichannel frequency-domain blind speech separation,” in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Orlando, FL, May 2002, vol. 1, pp. 881–884.
- [17] H. Saruwatari, T. Kawamura, T. Nishikawa, A. Lee, and K. Shikano, “Blind source separation based on a fast-convergence algorithm combining ICA and beamforming,” *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 14, no. 2, pp. 666–678, Mar. 2006.
- [18] J. M. Sanchis and J. J. Rieta, “Computational cost reduction using coincident boundary microphones for convolutive blind signal separation,” *Electron. Lett.*, vol. 41, no. 6, pp. 374–376, Mar. 2005.
- [19] H.-E. de Bree, W. F. Druyvesteyn, E. Berenschot, and M. Elwenspoek, “Three-dimensional sound intensity measurements using Microflow particle velocity sensors,” in *Proc. 12th IEEE Int. Conf. Micro Electro Mech. Syst.*, Orlando, FL, Jan. 1999, pp. 124–129.
- [20] J. Merimaa and V. Pulkki, “Spatial impulse response rendering I: Analysis and synthesis,” *J. Audio Eng. Soc.*, vol. 53, no. 12, pp. 1115–1127, Dec. 2005.
- [21] B. Günel, H. Hacıhabiboğlu, and A. M. Kondoç, “Wavelet-packet based passive analysis of sound fields using a coincident microphone array,” *Appl. Acoust.*, vol. 68, no. 7, pp. 778–796, Jul. 2007.
- [22] R. O. Schmidt, “Multiple emitter location and signal parameter estimation,” *IEEE Trans. Antennas Propag.*, vol. AP-34, no. 3, pp. 276–280, Mar. 1986.
- [23] P. G. Craven and M. A. Gerzon, “Coincident Microphone Simulation Covering Three Dimensional Space and Yielding Various Directional Outputs,” U.S. patent 4,042,779, 1977.
- [24] F. J. Fahy, *Sound Intensity*, 2nd ed. London, U.K.: E&FN SPON, 1995.
- [25] K. V. Mardia and P. Jupp, *Directional Statistics*. New York: Wiley, 1999.
- [26] J. P. Princen and A. Bradley, “Analysis/synthesis filter bank design based on time domain aliasing cancellation,” *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-34, no. 5, pp. 1153–1161, Oct. 1986.
- [27] S. A. Martucci, “Symmetric convolution and the discrete sine and cosine transform,” *IEEE Trans. Signal Process.*, vol. 42, no. 5, pp. 1038–1051, May 1994.
- [28] M. Bosi and R. E. Goldberg, *Introduction to Digital Audio Coding and Standards*. Norwell, MA: Kluwer, 2002.
- [29] M. R. Schroeder, “Integrated-impulse method measuring sound decay without using impulses,” *J. Acoust. Soc. Amer.*, vol. 66, no. 2, pp. 497–500, Aug. 1979.
- [30] Bang and Olufsen, “Music for Archimedes,” *CD 101*, 1992.
- [31] D. Schobben, K. Torrkola, and P. Smaragdis, “Evaluation of blind signal separation methods,” in *Proc. Int. Workshop Independent Compon. Anal. Blind Signal Separation*, Aussois, France, Jan. 1999, pp. 261–266.
- [32] K. E. Hild, D. Erdogmus, and J. C. Principe, “Experimental upper bound for the performance of convolutive source separation methods,” *IEEE Trans. Signal Process.*, vol. 54, no. 2, pp. 627–635, Feb. 2006.
- [33] G. W. Elko, S. L. Gay, and J. Benesty, Eds., “Super directional microphone arrays,” in *Acoustic Signal Processing for Telecommunication*. Norwell, MA: Kluwer, 2000, ch. 10, pp. 181–237.
- [34] O. Yilmaz and S. Rickard, “Blind separation of speech mixtures via time–frequency masking,” *IEEE Trans. Signal Process.*, vol. 52, no. 7, pp. 1830–1847, Jul. 2004.
- [35] H. Sawada, S. Araki, R. Mukai, and S. Makino, “Grouping separated frequency components by estimating propagation model parameters in frequency-domain blind source separation,” *IEEE Trans. Audio, Speech Lang. Process.*, vol. 15, no. 5, pp. 1592–1604, Jul. 2007.



**Banu Günel** (S’97–M’00) received the B.Sc. degree in electrical and electronic engineering from the Middle East Technical University, Ankara, Turkey, in 2000, the M.Sc. degree in communication systems and signal processing from the University of Bristol, Bristol, U.K., in 2001, and the Ph.D. degree in computer science from the Queen’s University of Belfast, Belfast, U.K., in 2004 for her work in audio and acoustical signal processing.

Since 2004, she has been a Research Fellow in Center for Communication Systems Research (CCSR), University of Surrey, Guildford, U.K. Her main research interests are array signal processing, spatial audio, acoustic imaging, psychoacoustics, and communication acoustics.

Dr. Günel is a member of the European Acoustics Association (EAA) and Audio Engineering Society (AES).



**Hüseyin Hacıhabiboğlu** (S'96–M'00) was born in Ankara, Turkey, in 1978. He received the B.Sc. (hons) in electrical and electronic engineering from Middle East Technical University, Ankara, Turkey, in 2000, the M.Sc. degree in electrical and electronic engineering from the University of Bristol, Bristol, U.K., in 2001, and the Ph.D. degree in computer science from Queen's University Belfast, Belfast, U.K., in 2004 for his research into the simplification of signal processing algorithms used in room auralization by the application of psychoacoustical knowledge.

Since 2004, he has been with the Multimedia and DSP Research Group (I-Lab), Center for Communication Systems Research (CCSR), University of Surrey, Guildford, U.K. His research interests include audio signal processing, room acoustics modeling and simulation, psychoacoustics of spatial hearing, and microphone array processing. He is one of the core members of the EPSRC-funded Noise Futures Network.

Dr. Hacıhabiboğlu is a member of the Audio Engineering Society and European Acoustics Association.



**Ahmet M. Kondoç** (M'91) was born in Cyprus. He received the B.Sc. (hons.) degree in engineering from the University of Greenwich, Greenwich, U.K., in 1983, the M.Sc. degree in telematics from the University of Essex, Essex, U.K., in 1984, and the Ph.D. degree in communication from the University of Surrey, Guildford, U.K., in 1986.

He became a Lecturer in 1988, a Reader in 1995, and then in 1996, a Professor in Multimedia Communication Systems and Deputy Director of the Center for Communication Systems Research (CCSR), University of Surrey, Guildford, U.K. He has over 250 publications, including two books on low-bit-rate speech coding and several book chapters, and seven patents. He has graduated more than 50 Ph.D. students in the areas of speech/image and signal processing and wireless multimedia communications, and has been a consultant for major wireless media terminal developers and manufacturers. Prof. Kondoç is also a Director of Mulsys, Ltd., a University of Surrey spin-off company marketing the world's first secure GSM communication system through the GSM voice channel.

Prof. Kondoç has been awarded several prizes, the most significant of which are The Royal Television Societies Communications Innovation Award and The IEE Benefactors Premium Award.